DOCUMENT RESUME

ED 255 284 PS 014 305

AUTHOR Hilliard, Asa G., III
TITLE The Future of Follow-Through.
PUB DATE [78]
NOTE 32p.
PUB TYPE Viewpoints (120)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Early Childhood Education; Educational Improvement; *Educational Quality; *Evaluation Needs; Program Development; *Program Evaluation; *Student Development; *Teacher Effectiveness; *Validity
IDENTIFIERS *Project Follow Through

ABSTRACT

Project Follow Through has not yet been properly evaluated by researchers guided by appropriate assumptions. Tests used in the basic Follow Through evaluation had no clear relationship to most Follow Through programs. Still, conclusions of the evaluation have been presented with certainty and lead to the judgment that Follow Through does not work. Suggesting otherwise is the fact that, at some Follow Through sites, between 10 and 13 percent of the children have scored above the 50th percentile on the Metropolitan Achievement Test. This fact should be examined in detail through comparisons of high- and low-performing programs. There is much to recommend such a comparison. Certainly, it is not difficult to find educational data that reveal long-lasting, positive outcomes of good teaching for students thought to have limited potential for achievement. Along with a focus on child outcomes, future Follow Through evaluations should focus on teacher behavior and the way treatments, or models, are implemented. It will be absolutely necessary to insure that measurement instruments employed are appropriate to the task. The methodology for making precise comparisons among educational treatments presently exists. Contrary to generally held views of experts, Follow Through has not been a failure; Follow Through children have not failed to respond to good teaching. When educators become clear about the goals of education and commit themselves and their resources fully to good teaching, academic success for children will be assured. The alternative is a dangerous academic rationalization and legitimization of the status quo. (RH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED255284

The Future of Follow-Through

Asa G. Hilliard III

Fuller E. Callaway Professor of Urban Education

Georgia State University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Asa G. Hilliard III

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

PS 014305

Large-scale educational evaluation studies have tended to conceal much more than they have revealed. This is particularly true of the major national studies on "program effects". Many of these studies, such as the Westinghouse Head Start Study (1969), the Coleman Study of Secondary Schools (1966), and the Abt Day Care Study (1977) are well known for their findings of small or no significant program effects. These repeated findings have been accompanied by deep-seated feelings of despair among professionals and parents who have devoted tremendous amounts of emergy toward the improvement of the quality of education for young children. However, their despair can in no way equal that which is found in communities which are served by such programs where the citizens have hoped for so much. If we take the results of the primary national studies of educational innovations at a national level in general, we might feel compelled to conclude that everything which can be tried has been tried and that, as some evaluators have suggested, we should lower our expectations, both for finding effective programs and for raising the achievement of children. If we are to believe the general evaluation results in education, then we must conclude that the children of the poor are doomed to an eternity of educational failure.

One exception to the type of results which are generally reported in major national studies of all types and at all levels of education is the work of Ron Edmonds (1979) and others which has attempted to study effective schools. The Edmonds approach to educational evaluation, or research, depending on your point of view, represents a radical and profound shift in the type of questions which the evaluator asks. The typical question in national evaluations of federal education intervention "programs" is: Do all schools, programs, classes, etc., on the <u>average</u>, work? Edmonds has turned the large scale

evaluation approach upside down. His question is at once both simple and profound. It is not a question of "does a national 'program' work?" but "do any school sites, and then, by extension, does any program work?" His somewhat startling finding is that there are indeed many "garden variety" schools that serve ordinary children, ordinary parents, in ordinary low-income neighborhoods, which do succeed in getting high quality academic results for children comparable to those which might be expected in other places, as far as the basic skills are concerned. In typical program evaluations, the nature of the aggregation of data has caused such instances of school success to be overlooked or submerged in the basic data. Edmonds' work was quite different. He took his analysis one step further. He chose to do a more fine-grained analysis of those schools which were successful in comparison to those schools which were quite unsuccessful, and was able to identify certain typical characteristics for successful schools. Then he took an additional, and most courageous, step, and went to work in the New York City Public School System to demonstrate that it would be possible, using data from his school effectiveness studies, to intervene in schools to improve success. The preliminary data from his New York City work indicates a significant degree of success in this regard.

To some extent, the Follow-Through evaluation is comparable to the Edmonds approach. On one level, it asks an Edmonds-type question: "Do any of the 'models' work?" However, at another level it repeats the errors of previous global studies of national education interventions, or of regular school operations. If the Follow-Through study had asked, "Do any Follow-Through sites work?" it undoubtedly would have located many such sites, as is suggested in parts of the Abt Report. However, by treating the "models" as if they represented unique and consistent treatments, it might well be expected that successful Follow-Through operations would be submerged in the aggre-

gated data, since few models are articulated precisely or guaranteed to have been implemented. More will be said about this later on.

When those who were responsible for making decisions about the development of a Follow-Through thrust were operating in the initial stage, at least two kinds of assumptions could have been made about the reasons for the poor performance of poor African-American children and other children served by the program: developers might have reasoned that the children's low achievement was due to the fact that they had been deprived of adequate instruction in school; or developers could have concluded that the poor performance of Follow-Through children was due to the fact that the children or their families possessed some deficit which required unique intervention strategies.

It makes all the difference in the world which of the two hypotheses designers happen to favor. If the former is accepted, then the Follow-Through program logically would simply be a matter of providing sufficient resources to parents in order that they might provide the same type of instruction which is available to the families which are more affluent than they. On the other hand, if the second hypothesis is accepted, then the Follow-Through program becomes a kind of experiment to test the validity of various unique "alternative" strategies which would solve the "puzzle" of how to teach the children who are served by Follow-Through. It appears that supporters and antagonists alike of Follow-Through programs are quite together, for the most part, in emphasizing the second assumption. It certainly appears to be reflected in the general character of the Follow-Through operation and in the character of the formal evaluations. For example, one might ask the hypothetical question, "if on the average, a parent of a Follow-Through child were provided with the resources to send their child to a high quality expensive private preschool and primary school program, would we expect there to be significant

positive differences for the child in the new setting versus the old?" To answer "yes" is to come down on the side of Follow-Through as a resource program, since program quality is knowable and determines the quality of a child's academic progress. To answer "no" is to come down on the side of Follow-Through as a problem-solving program, since there is a presumption that we do not know how to teach children who are poor. It should be remembered that the basis for the observations of Follow-Through programs in evaluation seem to derive from the second assumption.

The Validity of Evaluation Instruments

It is important that brief mention be made of the actual instruments which were used in the basic Follow-Through evaluation. Among those considered were the Wide-Range Achievement Test, Preschool Inventory, the Metropolitan Achievement Test, the Ravens Colored Progressive Matrices, the Coopersmith Self-Esteem Inventory, and a Questionnaire on Intellectual Achievement Responsibility. The four instruments which figured prominently in the final evaluation were the Metropolitan Achievement Test, the Ravens Colored Progressive Matrices, the Coopersmith Self-Esteem Inventory, and the Intellectual Achievement Responsibility Questionnaire. To anyone who is familiar with these instruments, it should be clear that their relationship to most Follow-Through programs is minimal to nonexistent. In this author's experience, few educators who use the results of measuring instruments are familiar with their contents. Content validity for measuring instruments is a basic professional requirement. For educators familiar both with Follow-Through programs and with the four instruments selected as criteria for use in program evaluation, the mismatch between programs and instruments for their evaluation should be crystal-clear. Of all the Follow-Through programs that I have visited, I have seen none which used Ravens Matrices as program content. Some psychologists and educators appear to support

the idea that the Ravens Colored Progressive Matrices Test is really a test of higher cognitive abilities rather than a simple achievement test. And yet, the work of Reuven Feuerstein in Israel has shown clearly that learners can be taught to solve Ravens type problems quite easily.

This leaves the Metropolitan Achievement as the only other test of academic achievement. Once again, it becomes critically important to determine empirically if the content of the Metropolitan Achievement test matches the content of Follow-Through programs. In the absence of such empirical demonstration, the use of the Metropolitan Achievement Test as a universal criterion for Follow-Through evaluation must be regarded as highly suspect.

In my opinion, the Coopersmith Self-Esteem Inventory is inappropriate for Follow-Through evaluation. Any program evaluation is dependent upon the use of valid instrumentation. And yet there are not sufficient data on the use of the Coopersmith Self Esteem Inventory with the diverse cultural groups represented in Follow-Through programs at the Follow-Through age to meet even the most minimal professional requirements. Little need be said at all about data in support of the Intellectual Achievement Responsibility Questionnaire. The basic point is this: without the existence of valid instrumentation, the program evaluation is fatally flawed. And yet when we come to final evaluation of the Follow-Through program, we find little admonitions for caution in the interpretation of the data.

It is hard to find a villain in this situation. Pressures are enormous on all parties to the Follow-Through experience. Any contractor who agrees to do an evaluation will be under extreme pressure to get the job done without excuses and with "the best instrument that we have at the time." Federal agency personnel likewise are under tremendous pressure to deliver to the Congress and to others definitive answers to the question "Does Follow-Through work?" Public policy makers also experience tremendous pressure to be account-

able for public funds and to demonstrate to constituents that effective action is being or has been made on their behalf. The combination of these and other pressures on these and other parties to the Follow-Through process forces evaluators to take short cuts. It is difficult if not impossible to avoid the expedient solution. And yet it can have dire consequences.

The Follow-Through Results

The Follow-Through evaluation has offered the following conclusions. Anderson, Pierre, Proper and Stebbins (1979) indicated that the following were their main findings:

1. "Each Follow-Through model had very different effects on test scores in the various communities in which it was implemented. Differences in effectiveness between sites within each model were greater than overall differences in effectiveness between models. None of the 17 models in the evaluation demonstrated that it could compensate consistently for the academic consequences of poverty. From this finding . . . with which the House group agrees, we conclude that the Follow-Through strategy of externally sponsored curricular change is not a reliable tool for raising the test scores of poor children; the strategy is sensitive to too many influences over which it has no control. Local circumstances and behavior clearly have more to do with children's test performance than do the intentions, theories, and rhetoric of outside interveners."

2. "In most cases the Follow-Through group scored about as one would expect similarly disadvantaged groups to score without Follow-Through. Where differences were apparent, Follow-Through groups scored lower more frequently than they scored higher. . . . it appears clear, then, that the Follow-Through strategy is not an effective tool for raising poor children's test scores. Not only are the effects unstable, but they are small on the average, and a disquietingly large minority of them are in the wrong direction."

3. "With few exceptions, Follow-Through groups were still scoring substantially below grade level at the end of three or four years' intervention . . . . poor children still tend to perform poorly in school, even after the best and the brightest theorists--with the help of parents, local educators, and federal funds, and supported by the full range of supplementary services associated with community action programs--have done their best to change the situation."

If the pessimistic conclusions which are presented above are taken at face value, the likely public policy consequences should be apparent to anyone. The basic point of these conclusions is that "Follow-Through does not work." The unfortunate thing about these conclusions is that they are presented with such a ring of certainty--certainty that the best efforts of a number of people have been given and that they have failed. And yet, what was the "Follow-Through program"? Was the program in all its varied forms, faithfully implemented? Were the evaluative criteria congruent with program goals? Did valid instrumentation exist to determine the extent to which the goals were achieved? To leave the matter as it stands is to level an implicit indictment of the children of Follow-Through and/or their families as being deficient in some ways. And yet there are at least two other sources of explanations for the apparent lack of progress among Follow-Through children. One the one hand, progress may have been achieved and yet not revealed because of invalid evaluation. On the other hand, progress may not have been achieved due to the quality of the educational treatment which Follow-Through children actually received. Up to now, a good deal of attention has been focused on the matter of the adequacy of the evaluation. Less attention has been focused on the precise nature and adequacy of the educational "treatment" which the children actually received. Because of this, the evaluation of Follow-Through has not yet been accomplished.

In "Perspectives on the Follow-Through Evaluation" (1978), with the exception of the broad view taken by Walter Hodges, critics and supporters of the Follow-Through evaluation study looked very narrowly at matters of experimental design and instrument validity. While discussants exhibited technical adequacy regarding the topics which they chose to review, to anyone who is familiar with the operation of Follow-Through operations on a day-by-day basis, these technical discussions would seem quite remote, not only because of their technicality, but because of the types of things which were not discussed but which are important in Follow-Through programming. Walter Hodges raises important questions, such as the matter of parent involvement, school climate, and curriculum development. All of these were important process goals for Follow-Through, yet they received little attention in the evaluation effort. But most important, Hodges points to the findings that at some Follow-Through sites, between 10 and 13 percent of the Follow-Through children were above the 50th percentile on the Metropolitan Achievement Test. There is a tendency, in the kind of analysis which has been performed in the evaluation and in the kind of analysis which serves as a basis for the critique, to treat such an unusual finding almost as if it were a random variation rather than as something which needs to be examined in detail. What was the nature of the Follow-Through program at the sites where the children's achievement was above the national average? High performing sites and low performing sites should be isolated for comparative analysis.

Because of the structure of Follow-Through historically, it would be difficult at this point to do such an analysis. For example, it is anything but clear that the Metropolitan Achievement Test reflected specific program goals for Follow-Through sites, even for those sites that did well on the test. And yet the Follow-Through program is too important to too many people

to allow such divergences in the trends to be submerged in national averages. Surely those teachers, parents and children at the high performing sites would feel devastated to learn that their achievements were effectively discounted. And yet these sites do not stand out in any of the materials on Follow-Through evaluation or critique.

A look at related experience. And so, given the data from formal evaluations, one could say that the great Follow-Through experiment seems to have taught us very little. This is not the same thing as saying that there were no lessons to be learned. Perhaps a different look at Follow-Through and alternatives with which it may be compared will suggest new ways of looking at this and at other such social programs in the future. It is hard to believe that anyone would question seriously the value of a quality education during the early years for young children which would be followed by attempts to sustain that level of quality instruction over additional years. In the minds of all parents who have the financial resources to cover the costs of education, there is no question but that education "works." We are left then with the question of whether Follow-Through is education, whether education works only for some children, or whether affluent parents are mistaken after all. In other words, is it less a question of whether the program works than whether some poor children can do the "work"? Perhaps there are still those among us who believe that children in poverty are so damaged by their conditions that they are unable to learn. On neither of these two questions does the formal Follow-Through evaluation provide us with adequate answers. And yet we may appeal to other data, should such an appeal be necessary. These data will show unequivocally that quality educational programs do "work" and that quality education works for virtually all children. If these things can be shown in other settings, then we have a new question: Why can't they be shown in the Follow-Through

formal evaluation data? I should be clear at once that by limiting the Follow-Through evaluation to a look at the Follow-Through program and to the Follow-Through population, it is highly likely that we have limited our ability to see important things about the education process in general and Follow-Through in particular.

There are many intervention or educational programs which achieve high-quality results with all types of children. Some of these are Follow-Through sites, as we have learned from Hodges' citation. The field of education is full of examples of dramatic achievement by students who were thought to belong to racial, ethnic or economic groups that had a limited potential for education. Some will be mentioned later. The power of teaching to change the predicted outcomes for such students has been demonstrated over and over again. The changes for students appear to be quite long-lasting. If the attention of educational researchers had been focused more closely on these successful examples, I do not doubt that there would be more abundant publicly known data to support this assertion. Even with the absence of a sufficient quality of long-term systematic evaluation attention, we may cite several examples. One of the most popular programs on 60 Minutes was a segment which featured West Side Prep High School, where Marva Collins is the Director. The children who are served at West Side Prep are children who live in one of the lowest socioeconomic areas in the city of Chicago. They are also children who, in most cases, were performing very poorly in school prior to the time of their entry into West Side Prep. The results of Marva Collins' dedicated work are there for all to see. The children have been overwhelmingly successful in precisely those areas of curriculum which were thought by many to be beyond their reach: mathematics, science, and, especially, English and literature.

A brief visit to the Oakland Community School in Oakland, California, reveals a school with similar results. Located in one of the worst poverty pockets in Oakland, it has shown children from, for the most part, very low-income families to be quite capable of mastering regular academic school subjects at the highest level. As in the case of West Side Prep, the children who have been helped are those who previously had had major difficulies in regular school environments. The same may be said for the Marcus Garvey School in Los Angeles, the Nairobi Day School in East Palo Alto, California, the McKinley Educational Institute in Berkeley, California, and a variety of other such schools scattered throughout the nation. We have already mentioned the work of Ronald Edmonds, which is important because his attention has been directed nationally toward ordinary "garden-variety" public schools which, against all odds, are succeeding beautifully. One wonders just how many successes are required before it is understood clearly that the quality of teaching still explains much more about pupil performance than does the socio-economic background of the students.

Our evidence for this is not limited to those few examples cited above, or even to the large-scale national study by Ron Edmonds. Reuven Feuerstein (1979) has shown clearly that "thinking skills" can be taught in a relatively brief period of time to thousands of "retarded" children who may have been thought to be uneducable beyond a minimal level. Renee Fuller (1977) has demonstrated the ease with which reading can be taught to "retarded" populations whose scores on IQ tests would indicate that they would be incapable of significant academic achievement. Paulo Freire (1973) has also shown that even as late as adulthood it is possible to teach reading in approximately 30 hours' time to adults who have spent all their lives in abject poverty. The list goes on and on. For those who would doubt the existence of such schools

and classrooms, or who would doubt the power of teaching to produce with comparative ease significantly greater levels of pupil achievement than we normally find in poverty areas, it is unlikely that extended citation of such examples of teaching power would serve to change those opinions. However, for those who do believe this evidence and who may have had many similar experiences of their own to report, a fundamental question arises: Why does the Follow-Through evaluation fail to yield information of the type which has just been cited above? If Follow-Through programs which were dramatically successful actually existed but were not revealed in the evaluation, then something is wrong with the evaluation. If, on the other hand, even the best Follow-Through programs actually yielded only minimal academic results, assuming the criteria and instruments to be appropriate, then something must be wrong with Follow-Through, since we do have the evidence that children's academic achievement can be changed, sometimes with quite minimal resources.

Some of the work in Head Start and Follow-Through, the research work, has been done for the purpose of gathering data to inform public policy decisions (Abt, 1977). Decisions about care-giver qualifications, group size, class size, and so forth, have been tied to information which is generated from a look at Head Start and Follow-Through programs. And yet, as has been shown, these national studies generally leave out the use of programs or schools which are known to be highly successful, even within the Follow-Through effort. Consequently, by habitually, arbitrarily, or thoughtlessly restricting evaluation attention to a comparison of Follow-Through "programs", a great opportunity has been lost.

Of course, this suggests some things for the future for Follow-Through research and evaluation. What precisely are the ingredients of the many successful early childhood programs and later childhood formal school exper-

iences which account for success in certain circumstances? If such research is conducted, one must be prepared for a fundamental challenge to some of our most cherished notions. Caregiver qualifications in successful schools bear little resemblance to proposed or publicized caregiver qualifications which are thought by many professional educators to be essential. At present, public policy decisions, if based on data at all, are based on the most limited data pool--data which come almost exclusively from federally sponsored programs. If Follow-Through evaluation gives us the opportunity to change that situation so that the wider arena in education is considered, a major contribution will have been made to evaluation and to the field of pedagogy.

I for one remain unconvinced by the formal Follow-Through evaluation data that the true power of the Follow-Through effort has been minimal. I believe that the evaluation approach itself is much weaker than the Follow-Through effort which it seeks to assess. Neither articulated criteria nor measuring instruments of sufficient sensitivity exist at the present time to allow us to do valid assessment of Follow-Through. Further, if all Follow-Through programs are expected to have the same goals and objectives, then that policy decision must be reflected in program design, staffing, implementation, and supervision. For example, we know enough now to design educational experiences for children which will cause almost all of them to meet the minimal academic competencies which schools require. In fact, for most if not all children, achievement should be well beyond the minimum academic requirements. While formal data from the Follow-Through experience may not have taught us this, the fact that it has not suggests a need for a major overhaul of evaluation, program or both. These are policy matters. We cannot have planned variation in programs and planned restriction in

program outcomes. Either policy decision must be made to design the evaluation in advance of Follow-Through programs and to link program planning to the explicit goals, or the program evaluation should be eliminated altogether, as it is with education generally.

It is my opinion that evaluation does not serve us well if it sets up a mental orientation that allows for the acceptance of conclusions from present Follow-Through evaluation data, along with the belief that "Everything that could be done has been done." In the final analysis, Follow-Through children need one thing more than anything else: they need to be guaranteed a high quality education. It is only after that quality has been guaranteed without the expected success in children's achievement that special "interventions" or research need be conducted as a way of helping children.

And so, in the case of Follow-Through, the Office of Education did not offer its support to programs of known quality, ask that they be implemented and that their results be "evaluated"; it asked essentially a set of research questions which may be summed up as follows: "Can the problem of finding suitable ways of educating the poor be solved?" "The hope of many individuals in Washington was to find better ways to educate poor and minority children" (Harvard Education Review, 1978, p. 125). To ask if an educational program "works" prior to the time that either the program or its objectives can be articulated is to ask a research question which uses the language of evaluation. Follow-Through research data has been treated as if it were evaluative data. We should have learned by now to distinguish between these two things.

The failure to make this important distinction has major consequences for communication and for the mobilization of support for a social program among the various program constituents. It is important to have agreement on who is responsible for what and to know in advance just what a "program" can

or cannot do. If parents and children are participants in an experiment, then they should be so informed and the experiment must fall under the informed consent guidelines for the protection of human subjects. On the other hand, if the details of a non-experimental program of known quality are described fully, then all parties to the program can and should be in a position to monitor every aspect of its implementation. Then evaluation is possible.

Defining a Follow-Through model. A close inspection of the descriptions of Follow-Through models will reveal that there are insufficient data to determine with adequate precision just what the unique character of the educational model was in virtually any of the cases. This certainly is not to say that good education could not have been conducted, or was not conducted, under the rubric of a "model." In fact, in the judgment of many, good education did occur at many Follow-Through sites. However, an examination of descriptions of Follow-Through models fails to show that a paradigm exists for model definition which would allow precision distinctions to be made among the variations for either research or evaluation purposes.

Keep data on the implementation of a "model". Assuming that the particulars of a model could be articulated according to some common way of talking about models, another step would have to be taken: that is, to determine if the model which had been described for a given Follow-Through program was uniformly implemented at all sites throughout the program. This point was made in the House, Glass, McClean and Walker critique (1978). The failure to do this has resulted in the apparent presumption that the loosely described "models" were in fact uniformly and fully implemented, since there are few data in the technical reports which allow for a judgment to be made regarding the degree to which the stated model was implemented. Indeed, it would be a major evaluation effort simply to verify that implementation of models.

> Stallings' initial problem was specification. There were no comprehensive descriptions of sponsors' models in terms of classroom behavior. Without such descriptions it would be impossible to assess how well models had been implemented. So in the spring of 1973, Stallings visited "ideal first and third grade classrooms specified by each sponsor" in the hope of defining models in behavioral terms by observing ideal model classrooms. "What would be desirable for a closely controlled experiment would be an explicit statement by sponsors of what proportion of the time critical variables would occur in an ideally implemented classroom. However, elements in a classroom are not like those in a test tube. We have not yet learned to predict the amount of individual attention or feedback that a group of children needs to meet specific goals. Each classroom is made up of individuals, and individuals are most likely to need different rates of feedback or individualized instruction for maximum growth. Thus, even the most specific models, such as the University of Oregon and the University of Kansas, would alter the rate of reinforcement depending upon the needs of the child. . . . We simply could not come up with a recommendation that would specify the rate of questions or reinforcement to be expected from even the most structured models." In consequence, the attempt to specify ideal model types directly did not work. (The Follow-Through Report, Vol. V, p. 157)

The failure to define a model adequately and to guarantee its implementation in an evaluation study has placed Follow-Through children and their families at great risk, since, as data are reported, the public and casually informed policy makers assume that articulated programs were fully operational. Therefore the explanation for failures to find large significant differences between Follow-Through children and non-Follow-Through children tends to locate the blame for "program failure" with the children and/or their families. One could predict with a fair degree of certainty that the next step to help Follow-Through children would in all likelihood not be attempts to look at models, but rather to devise new "models" for the purpose of solving the continuing "puzzle" of how to educate children who have been supported by Follow-Through. The question always seems to come out, "does education work?" We never seem to be able to ask the question, "did education happen?" or "what type of schooling happened?"

> It is worth noting again that among all the mounds of material resulting from the Follow-Through evaluations, these descriptions by Stallings are unique. They comprise the only direct empirical descriptions of what life is like in the classrooms of different sponsors. (The Follow-Through Report, Vol. V, p. 159)

One of the most important lessons from the Follow-Through program is that the treatments which children receive vary. Perhaps an unintended consequence for the conduct of Follow-Through evaluations is to document the variety of treatment which children receive both within and outside school, or rather, treatments in special programs when compared to non-programs. Many professionals appear to hold an image in their heads of teaching as a standard "treatment".

> All five of the major Follow-Through studies--the Emerich and the Abt I, II, III, and IV Reports--discuss the implementation issue and its importance to interpreting effects estimates, but none of these Follow-Through evaluators, neither Emerich et al., nor Klein and company, nor Anderson, Stebbins and associates, could make much progress on the problem. Their reactions to this frustration were similar--they wanted more information. Emerich proclaimed the "need for more precise treatment definitions and descriptions" . . . Klein and company called for a "directory of effects associated with types of sponsors under different types of conditions" . . . and in their summary of the Abt IV Report, Stebbins and associates advised "future evaluations that aspire to explain the outcome of compensatory intervention will do well to investigate thoroughly the local circumstances, attitudes and activities that clearly decide whether a compensatory education program can work or not. (The Follow-Through Report, Vol. V, p. 157)

It has been only in recent years, through the work of people such as Jane Stallings at Stanford Research Institute, Fred McDonald at the Educational Testing Service, Jeri Brophy at the University of Texas, Raymond Rist, and others, that the wide variations in treatment which children receive have been described systematically. The effect of these data should be to cut the ground from under those who choose to attribute to children or to their families the sole or the major responsibility for children's school achievement or the lack of it. One of the hardest things to do in

education has been to focus professional attention on what professionals do. Even in the Follow-Through evaluation, when the data are provided on educator behavior, it appears to be an afterthought comprising the smallest part of the study; and while, in my opinion, it should form a major and indispensable part of any interpretation, it is virtually absent from the final summary report of the evaluation. Even then it occupies too little of the attention of the critics. Weikert and Benet have summed the matter up nicely: "It is not necessarily the methodology that is at fault, even in 'quasi-experiments' like Follow-Through . . . what is needed may be a clearer conceptualization of models, outcomes, and intervening processes, as well as better educational treatment for the researcher to evaluate" (The Follow-Through Report, Vol. V, p. 164).

Instrumentation. The use of an achievement test, the MAT, an IQ test, the Coloured Ravens Progressive Matrices, a self-concept test, the Coopersmith Self Esteem Inventory, and a home-made "locus of control" test as the criteria for whether a program works or not, is much more than a necessary compromise in order to utilize the "best instruments available." It is a virtual surrender to triviality. There simply must be some point beyond which essential meaning for evaluation has been compromised away. If existing standardized instruments allowed most of our evaluation questions to be answered, one might say that a compromise which settled for incomplete or inadequate instrumentation would be useful. On the other hand, when the mismatch between the content of instruments and program descriptions is as gross as appears to be the case in Follow-Through, especially with "planned variations", then a little bit of information may well be worse than none at all. For example, to use the Raven Progressive Matrices Test as the measure

of Follow-Through outcomes is as big a mistake as can possibly be made. On the one hand, if the Raven Matrices Test is thought of as a test of "intelligence", one would not expect changes in such scores as a consequence of Follow-Through. On the other hand, we do have data that highly effective strategies for solving Raven type problems can indeed by taught to learners, causing dramatic changes in the scores that they earn on such tests (Feuerstein, 1979, 1980). If that is the case, then a large part of what the Raven measures must be achievement. Therefore it would be inexcusable, in the absence of such an explicit achievement goal, for a given Follow-Through model to be evaluated utilizing this achievement test, which has no demonstrated content validity for any programs which are being assessed. It seems to be clear that the basic Follow-Through evaluation required major compromise. It was compromised in the direction of setting up the study to fit resources, instrumentation, and unintended types of field implementations. By the time all those compromises had been made, very little could be said from the evaluation about a "Follow-Through program". If all Follow-Through programs were to be evaluated by the same narrow and, in my opinion, minimally relevant, criteria, program operators should at the very least have had advance warning. They should have been provided with the instruments so that they might adjust their programs to coincide with such externally imposed objectives. We should have learned of the necessity for the development or selection of appropriate valid instruments for evaluation. Too much is at stake to permit casual attention to this matter. If it is important to ask either research or evaluation questions, it is important that equal concern be reflected for the provision of measures appropriate to the task. It is not enough to "clean out the cupboard" and use leftover tests, since it is not only programs but their clients who are evaluated ultimately.

What is not revealed. As one reads the materials from a variety of Follow-Through evaluations, one is struck by how cold, remote and irrelevant to actual program operation the results of these evaluations appear to be. Because of the constraints of evaluation procedures, the results yield little that will help readers to grasp a real flavor of what Follow-Through is all about. While important at one level, the debate between evaluators and their critics resembles more of a scholastic exercise than a vital function for educational programs. One is left with the sinking feeling that no matter who might win in the debates over units of analysis or the use of analysis of covariants as an appropriate analytical tool, little would have been learned from the past evaluations to reveal the dynamic workings of the successful educational process itself or basic positive results which it might indeed have achieved. Therefore, in once sense, the lesson of Follow-Through is more a lesson about what to do and what not to do in program evaluation than it is about how to do Follow-Through education or whether Follow-Through education worked.

From what has been learned from the operation and evaluation of Follow-Through programs to date, it is not possible to offer a bold redesign of education for children who are served by the program. We are faced with an uncomfortable situation. On the surface, it would appear that we can do just as much for Follow-Through children by leaving them completely alone as by involving them in special programs. As indicated earlier, I am unable to accept such a conclusion for several reasons:

1. Follow-Through "programs" have yet to be described in terms which are appropriate for rigorous evaluation.
2. Content-valid instrumentation for all Follow-Through program evaluation does not exist at this time.

3. There is a need to be clear about the rationale for a Follow-Through program. Is Follow-Through needed as an experiment to determine how to solve the problem of how to educate poor children? Or, is Follow-Through to be an attempt to provide the resources so that children who are now deprived of normal school experiences can be given them?

4. There are many demonstrations which prove the effectiveness of good teaching.

To date, only the population that is or could be eligible for Follow-Through has been a part of evaluation efforts. Secondly, even in those cases, there has been insufficient documentation to provide a description of the actual versus the promised treatment of Follow-Through children among the many models or between the educational treatment of those who are provided a Follow-Through treatment and those comparison groups who were not a part of the treatment. But most important of all, there is no documentation to compare the actual treatment which is offered to Follow-Through children and children in non-Follow-Through programs where children experience high rates of success in school achievement.

As indicated earlier, the methodology for making precision comparisons among educational treatments already exists. Jane Stallings at Sanford Research Institute, J. Brophy at the University of Texas, David Berliner, formerly of the Far West Regional Laboratory for Educational Research and Development, among others, have developed observational methods for classrooms. Such methods can be used both to verify the implementation of promised programs, and to determine the nature of such distinctions as may exist among programs and between programs and regular good educational offerings. Up to now, the emphasis of evaluations has been on child outcomes. In the future, in the interest of valid evaluation and in fairness to Follow-Through children, both child outcomes and school treatment must be considered simultaneously.

The National Institute of Education is in an excellent position to structure a new research and evaluation effort. In view of existing Follow-Through evaluation data, it is urgent that such an effort be initiated and that it be comprehensive. The low expectations for school achievement which many educators seem to have for children like those who are served by Follow-Through programs are fed by research and evaluation programs which are underfunded and therefore incomplete and invalid, limited in conception and therefore incomplete and invalid, or restricted to Follow-Through children and therefore incomplete and invalid.

The Basic Lessons of Follow-Through

There are a number things which can and should become lessons for educators from the Follow-Through experience. They are as follows.

Stop confusing research and evaluation. When the Office of Education asks the question, "Does Follow-Through work?" it sounds as if it is an evaluation question. In reality, it is not. It is a research question, for the simple reason that, at the outset, neither Follow-Through itself nor the Planned Variations were programs of known instructional quality. Evaluation implies the implementation of knowns. Neither, at the outset, were there testing and evaluation instruments which were designed specifically for the purpose of evaluating a full range of Follow-Through programmatic outcomes; indeed, the state of the art in testing and measurement is such that the very process of designing valid instruments may well be as large as the Follow-Through operation in total. Further, the general goals of Follow-Through were not then, nor are they now, restricted and articulated in a way that would allow a precision evaluation to proceed. In this sense, Follow-Through is not unlike "regular education". In fact, until the social programs of the 60's were introduced, it was rare that educators were expected to prove that education "worked" as a price for fiscal support.

Clarify Follow-Through purposes. At a public policy level, it is necessary to determine clearly if Follow-Through is to be a support or an experimental program. Whichever is to be the case, the policy position must be supported with the best possible professional rationale. That rationale should be grounded in actual educational experience and not in futuristic speculation. In other words, if Follow-Through is to be support for a program of known quality, examples of successfully operating programs which serve as models for public policy must be available. If the Follow-Through program is to be seen as a research program, then it is most unfair to Follow-Through children to implement it on a national scale.

Free Follow-Through from the state of the present art in evaluation. The plain truth is that professional judgment may be a better way of determining the value of the Follow-Through experience for children than the mechanism of traditional program evaluation. The state of the art in instrumentation for program evaluation is not sufficiently developed to allow the future of Follow-Through programs to ride on traditional program evaluation outcomes, especially since the state of the art in educational programming is not so well developed that a uniform Follow-Through experience can be mandated at a national level, nor should it. Decisions about the support of Follow-Through must be made long before the state of the art in evaluation can be improved substantially. Perhaps a combination of client satisfaction and professional judgment is the best that we can hope for on the evaluation side at the present time.

Build both a research and evaluation capacity. My concerns and criticisms with existing evaluation and research are not intended in any way to suggest that either evaluation or research in educational programs such as Follow-Through should be eliminated. Quite to the contrary, both evaluation and research have the potential for making major contributions to the field of educa-

tion. However, this cannot happen unless sufficient fiscal support is offered for that purpose. The demands and expectations of sponsors of research and evaluation must be met by an appropriate fiscal level, but above all, educational researchers must be given the time to do the job. The needs of public policy makers for adequate information may be strong, yet educators must not be asked to do what cannot be done.

Conclusion

It has been hard for many educators to realize that superior school achievement is associated with superior teaching. The belief persists among many professionals and among influential members of the lay public that families are more responsible for school achievement thatn are schools, overlooking the fact that families who have the resources are able to guarantee access to the best schools for their children. There is little empirical research to document this phenomenon. However, it takes little observation in schools which serve the children of the wealthy and in schools which serve children of the poor to see the dramatic differences between them in educational experiences which they provide for children. The victims of the pervasive belief that "school experience" has minimal effect on children are the children who are served by Follow-Through programs. It is the continuing failure of educators to distinguish among the variety of qualities of school experiences which allows "school experience" to be perceived as a common thing. It is this which allows the failure of educational systems to be seen as failures of children and families.

Follow-Through has not been a failure. I believe that we may say that it has succeeded in direct proportion to its provision of appropriate resources to support good teaching. Follow-Through children have not failed to respond to good teaching. It remains for educational leaders and policy makers to stop the drift in thinking which leads to a loss of faith in children and edu-

cators. When we become clear about the goals of education, and when we commit ou selves and our resources fully to those goals, the academic success for children will be assured. The alternative is a dangerous academic rationalization and legitimation of the status quo.

References

Abelson, W. D. and others. Effects of a four-year Follow-Through program on economically disadvantaged children. Journal of Educational Psychology, October 1974, 66 (5), 756-771.

Abt Associates, Inc. Education as experimentation: A planned variation model (Volume IV-C, appendices, parts I and II). Cambridge, Massachusetts, 1977.

Bond, J. T., Smith, A. G., & Kittel, J. M. Evaluation of curriculum implementation and child outcomes (Annual report, Cognitively Oriented Curriculum Project Follow-Through). Ypsilanti, Michigan: High/Scope Foundation, 1976.

Califano, J. A., Jr. The appropriateness of the Federal Interagency Day Care Requirements (FIDCR): Report of findings and recommendations. U. S. Department of Health, Education and Welfare, Office of the Assistant Secretary for Planning and Evaluation, Washington, D. C., 1978. U. S. Government Printing Office, Stock Number 017000002112.

Cline, M. G., Ames, N., Anderson, R., Bales, R., Ferb, T., Joshi, M., Kane, M., Larson, F. Park, D., Proper, E. Stebbins, L., & Stern, C. Final report: Education as experimentation, evaluation of the Follow-Through planned variation model (Volume IA: Early effects of Follow-Through and Volume IB: Monographs). Cambridge, Massachusetts: Abt Associates, 1974.

Cline, M. G., Ames, N., Anderson, R., Conte, M., Ferb, T., Hatt, C., Joshi, M., Larson, J., Park, D., Proper, E., Stebbins, L., & Stern, C. Final report: Education as experimentation, evaluation of the Follow-Through planned variation model (Volume IIA: Two-year effects of Follow-Through

Coleman, J. S. et al. Equality of educational opportunity. Washington, D. C.: U. S. Government Printing Office, 1966.

Edmonds, Ron. Some schools work and more can. Social Policy, March/April 1979, 29-32.

Freire, P. Education for critical consciousness. New York: Seabury, 1973.

Fuller, R. In search of the IQ correlation: A scientific whodunit. Stonybrook, N. Y.: Ball-Stick-Bird, Inc., 1977.

Haney, W. The Follow-Through planned variation experiment (Volume V: A technical history of the national Follow-Through evaluation). Cambridge, Massachusetts: Huron Institute, 1977. Department of HEW Contract # OEC-0-74-0394.

Haney, W. Draft report-reanalysis of Follow-Through parent and teacher data. Cambridge, Massachusetts: Huron Institute, 1978.

High/Scope Educational Research Foundation. Annual report: Fifth Follow-Through sponsors conference. Ypsilanti, Michigan: Author, 1975.

Hodges, W. L. The worth of the Follow-Through experience. Harvard Educational Review, 1978, 48, 186-192.

Hodges, W. L., & Sheehan, R. Follow-Through as ten years of experimentation: What have we learned? Young Children, 1978, 34, 4-18.

House, E. R., Glass, G. V., McLean, L. D., & Walker, D. F. No simple answer: Critique of the "Follow-Through" evaluation. Harvard Educational Review, 1978, 48, 128-160.

Kennedy, M. M. Findings from the Follow-Through planned variation study. Educational Researcher, 1978, 7, 3-11.

Lazar, I., Hubbell, V. R., Murray, H., Rosche, M., & Royce, J. The persistence of preschool effects: A long-term follow-up of fourteen infant and preschool experiments (Final Report, Grant No. 18-76-07843 to the Administration on Children, Youth and Families, Office of Human Development Services, DHEW). Ithaca, New York: Cornell University, 1977.

McNamara, T. Follow-Through pupil achievement characteristics in Philadelphia: Quasi-longitudinal data, 1975-1976 and 1976-1977 (Report No. 7870). Washington, D. C.: Office of Education (DHEW), 1978.

Mann, A. J., Harrel, A., & Hurt, M., Jr. A review of Head Start research since 1969 and an annotated bibliography. Department of Health, Education and Welfare Contract HEW 105761120. DHEW Publication # (OHDS) 78-31102.

Maraschiello. R. Post-program achievement of prekindergarten children in Philadelphia: Quasi-longitudinal data, 1974-1975 through 1976-1977. (Report # 7901).

Roupp, R., Travers, J., Glantz, F., Coelen, C., & Smith, A. Children at the Center: Final report of the National Day Care Study, summary findings and their implications, Vol. I. Administration for Children, Youth and Families, Office of Human Development Services, Contract # 105741100. Cambridge, Massachusetts: Abt Associates.

Seitz, V., and others. Long-term effects of intervention: A longitudinal investigation. National Institute of Mental Health (DHEW), Rockville, Maryland, Office of Child Development (DHEW), Washington, D. C., 1976.

Soar, R. S. Final report: Follow-Through classroom process measurement and pupil growth (1970-71). Gainesville, Florida: University of Florida College of Education, 1973.

Stallings, J. A. Follow-Through program classroom observation evaluation, 1971-1972. Menlo Park, California: Stanford Research Institute, 1973.

Stallings, J. A. Implementation and child effects of teaching practices in Follow-Through classrooms. Monographs of the Society for Research in Child Development, 1975, 40.

Stebbins, L. B., Anderson, R., Baltzel, C., Brooks, C., Evans. P., Ferb, K., Ferb, T., Freyheit., R., Friedman, S., Hagerty, M., Jastrzab, J., Katzenstein, E., Keefe, K., Ladner, R., Mackiernan, Y., Mingolla, E., Nelson, B., Piper, C., Poppe, C., Proper, E., St. Pierre, R., & Watkins, M. Education as experimentation : A planned variation model (Vol. 2-A: Findings, Cohort II and Interim Findings, Cohort III). Cambridge, Massachusetts: Abt Associates, 1976.

Stebbins, L. B., St. Pierre, R. G., Proper, E. C., Anderson, R. B., and Cerva, T. R. Education as experimentation : A planned variation model (Vol. 4-A: An evaluation of Follow-Through). Cambridge, Massachusetts: Abt Associates, 1977.

Weikart, D. P., Epstein, A. S., Schweinhart, L., & Bond, J. T. The Ypsilanti Preschool Curriculum Demonstration Project: Preschool years and longitudinal results. Ypsilanti, Michigan: High/Scope Foundation, 1978.

Weisberg, H. R., & Haney, W. Longitudinal evaluation of Head Start planned evaluation and Follow-Through. Cambridge, Massachusetts: Huron Institute, 1977.

Westinghouse Learning Corporation & Ohio University. The impact of Head Start: An evaluation of the effects of Head Start on children's cognitive and affective development. Washington, D. C.: Office of Economic Opportunity, 1969. (Distributed by Department of Commerce, Clearing-

house for Federal Scientific and Technical Information, Springfield, Va.)

Villaume, J., & Haney, W. The Follow-Through planned variation experiment (Vol. 5, Appendix: Analysis of interim Follow-Through evaluation reports). Cambridge, Massachusetts: Huron Institute, U. S. Office of Education, Department of Health, Education and Welfare, Contract No. OEC-0-74-0394, 1977.